ICR Workspace Working Group Teleconference Meeting Minutes

Date, Time & Location:	May 12, 2004, 2:00pm – 3:00pm EDT
Attendees:	Kristina Vuori – Burnham Kutbuddin Doctor – Burnham David Jewell - Dartmouth Patrick McConnell – Duke Jennifer Shoemaker - Duke Cathy Wu - Georgetown Craig Beam – Moffitt Steve Eschrich - Moffitt Edith Zang - The Institute for Cancer Prevention Mathieu Wiepert – Mayo Clinic Gary Bader – Sloan Alex Lash - Sloan Judith Goldberg – New York Edwin Quick – Oregon Health Vincent Vuori - Oregon Health Jack London – Thomas Jefferson Devjani Chatterjee - Thomas Jefferson Ajay Jain – UC San Francisco Jay Lewis – U of C Marsha Rosner – U of C Tom Casavant - Holden Terry Braun – Holden George Wu - Lineberger David Fenstermacher - Penn James Lyons-Weiler - Pittsburgh Mary Edgerton - Vanderbilt Rakesh Nagarajan- Wash U Harold Reithmans – Wistar Louise Showe - Wistar Margaret Borwhat – Patient Advocate John Powell – NCI Subha Madhavan – NCI Leslie Derr – NCI Lynette Grouse – NCI Carl Schaefer - NCI David Kane – SRA Weichou Leung – Tulane Paul Spellman – Lawrence Berkley Lab Claire Zhu – BAH Juli Klemm - BAH
Review workspace activities:	**Communications among SIGs within the workspace** ▪ Notes will be first shared with SIG participants for review, then posted to the caBIG online forum. ▪ Updates will be given at the monthly ICR WS/WG meeting. ▪ Regular meeting schedules have been established and will be distributed to participants. SIG mailing lists will be based on

attendance and requests.

- Mission Statement is being drafted by each team, and will be communicated to the rest of ICR within the next few weeks.
- Computational Genomics SIG is being rolled into Gene Annotation SIG.

Highlight of each SIG meeting

- Translational
  - Focused on tools, technologies and data that are necessary to integrate clinical data with experimental data.
  - The group discussed the importance of experimental design and agreed that providing tools to guide this process should be a long-term goal of this group.
  - A short-term goal is to produce a points-to-consider white paper to provide guidance on experimental design.
  - The group also discussed controlled vocabularies of interest, and will put together a list to be communicated to the Controlled Vocabularies and Common Data Elements group.
- Pathways
  - Support basic research and ICR tools development by providing easy access to pathway data and commonly used pathway analysis tools.
  - Develop common APIs and object models for pathway data to enable easy integration.
  - Develop common format for pathway information to allow easy sharing.
  - Provide data in a non-redundant format.
  - Develop pathway curation tools.
  - Provide more signaling pathways and other cancer relevant data.
  - Sloan has an extensive list of pathway data resources. The group will add to this and make it available on the caBIG website.
- Microarray Repositories
  - Identify and prioritize the needs for capture, storage and utilization of microarray data and related genetic data.
  - The group discussed the storage of data as flat-files vs. relationally. Flat-files facilitate interaction with downstream analysis tools; storing data relationally can enable powerful queries.
  - There is a need for a robust repository to organize and archive microarray data.
  - The group discussed the MIAME standard.

- While standards are good, providing this information can be burdensome if not captured along the way.
- Are there ways to incent experimental scientists to capture this information?
- It is important that such standards not get in the way of getting work done. It may not be necessary for some experiments such as initial pilot studies.

- Many groups are interested in NCICB's caArray that will be released in September. There will be on-line presentations on May 17 and June 2.

- Gene Annotation
  - Concerned with data and tools relevant to describing genes and gene products.
  - The group discussed the scope of the term "Gene Annotation", which ranges from IDs and names to functional descriptions and genomic features.
  - The group discussed which gene/gene product identifiers were most commonly used. Generally, Refseq IDs, LocusLink IDs, and Ensemble were cited as most reliable.
  - There is often ambiguity in annotation assignment. The group discussed how this should be handled. Providing a flag to the users and providing all available information to allow the user to sort out the discrepancy was cited as the most common approach.
  - GO is widely used, but it is important to capture and be mindful of the evidence code for the term assignment so as not to be misled.
  - The Computational Genomics SIG had a small number of participants and it was agreed this SIG should be rolled into the Gene Annotation SIG.
- Data Analysis and Statistical Methods
  - Focused on tools and algorithms to support high-level data interpretation.
  - The group had a lively discussion of how to support the experimental scientists' struggle with statistical data analysis.
    - Agreed that there is a shortage of biostatisticians to help at all stages of an experiment – often referring to microarray data in this context.
    - Many or most experimental scientists do not have the training to do the appropriate data analysis.
    - Many of the current tools are not user friendly and/or may lead to the generation of misleading results.

	• User training, improved documentation and good sample datasets were highlighted as being crucial. • This will be an ongoing conversation within this SIG. The group will likely put together a position paper on this issue. o Informatics for proteomics • Focused on tools and technologies to support the capture and analysis of proteomics data. Currently encompasses all platforms relevant to the study of proteins. • Technologies relevant to the attendees were 2D gels, MALDI-TOF, SELDI-TOF, and protein arrays. • Most studies focused on protein identification and quantification for the purpose of biomarker identification. • Standards for proteomics data are just emerging. MIAPE was called out as possibly leading the way. One member of this SIG will be involved in this effort and would be a de-facto caBIG liaison. • Integrating proteomics data with other types of data is of interest and being addressed at some of the centers.
Update from other Workspaces:	▪ Architecture (Liaison: Patrick McConnell, Duke) o Mission statement being generated o Goal is to develop a ubiquitous data computation grid that is accessible across the world. o Multiple subgroups have been formed: ▪ Interface architecture ▪ Information architecture => modeling formalisms, data mapping, universal ID mechanisms ▪ System architecture => data and computational grids, workflows, messaging ▪ Security and access control ▪ Software development => best practices, documentation ▪ Reference implementations ▪ Extended group: Communication with external groups outside of caBIG ▪ Controlled Vocabularies and Common Data Elements (Liaison: Mary Edgerton, Vanderbilt) o Developing a model for how CDEs will be governed across the community. Recommending that CDEs be created by scientists and be reviewed by the Controlled Vocabularies

	and Common Data Elements WS along with NCI. ▪ Tissue Banks and Pathology Tools (Liaison: Rakesh Nagarajan, Wash U) o Working on mission statement o Use Case template has been created that will be used by the Developers and Adopters in this Workspace. ▪ Clinical Trials Management (David Fenstermacher, Penn) o WS has created a Facts sheet outlining its charter/ o Reporting mechanisms – especially adverse-event reporting – were seen as high-priority o The C3D application is hosted by NCI as a tool for managing and hosting clinical trials. The backend is Oracle Clinical but a number of other clinical trials applications are working on caBIG compatibility. o Video/audio training sessions are ongoing. ▪ Training (Liaison: Edith Zang, Institute for Cancer Prevention) o Objectives identified: Internal communication, External communication, Training. o Defined target groups for training: developers, adopters, users. Subcommittees have been formed around these groups. o Determined that one of the sub-committees of the Architecture group, "software best practices" was also involved with training. o Communications plan is under development o Face-to-face meeting of this working group is scheduled for June 28-29 in Washington D.C. ▪ Strategic Planning (Communicated by Tom Casavant) o Group has focused on two general areas ▪ Communication methods: Two pilots have been launched to test protocols for video and data sharing ▪ Overall strategic planning for caBIG ▪ Data Sharing and Intellectual Capital (Liaison: Tom Casavant, Terry Braun, Holden) o Focused on privacy and intellectual property issues. Sense that these issues need to be resolved soon to facilitate caBIG projects
Update on Statement of Work process and progress:	▪ One-on-one conversations held with most of the Developers and Adopters to discuss project of interest. Proposals requested from developers based these conversations.

	▪ SIG meetings have provided a forum to begin the process of Developer/Adopter pairing. ▪ Statement of Work templates have been created for Developers, Adopters and Participants. Much of the content will be the same for all centers but the specific project activities will tailored to each project. This will be developed in close collaboration with the Centers. ▪ SOW creation is happening in parallel with the contract negotiations. The goal is that when the contract is signed off, a Statement of Work will be in place to allow work to begin immediately. All Centers are at different stages of negotiation -- it is a good idea to find out the status of the contract for your Center.
General caBIG announcements	▪ The caBIG website has been updated with URLs to the tools from centers in this group that are currently available: **http://cabig.nci.nih.gov/inventory/ICR_Tools** ▪ There are several upcoming webcast presentations that are of interest listed on the caBIG calendar: **http://cabig.nci.nih.gov/caBIG/calendar** For those who cannot attend these, they are being recorded and will be made available online after the presentation. ▪ If not already subscribed to the caAnnounce mailing list, please do so at **http://list.nih.gov/cgi-bin/wa?SUBED1=cabig_announce&A=1**

Action Items:

Name Responsible	Action Item	Date Due	Notes
Juli Klemm	Write and distribute meeting minutes	6/4/04	
Juli Klemm	Distribute dates and information regarding upcoming caArray presentations	5/14/04	